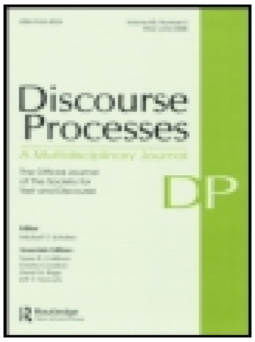

Discourse Processes

Routledge
Taylor & Francis Group

ISSN: 0163-853X (Print) 1532-6950 (Online) Journal homepage: http://www.tandfonline.com/loi/hdsp20

Cumulative Repetition Effects Across Multiple Readings of a Word: Evidence From Eye Movements

Juan E. Kamienkowski, M. Julia Carbajal, Bruno Bianchi, Mariano Sigman & Diego E. Shalom

DISCOURSE PROCESSES
http://dx.doi.org/10.1080/0163853X.2016.1234872

Cumulative Repetition Effects Across Multiple Readings of a Word: Evidence From Eye Movements

Juan E. Kamienkowski[a,d,e], M. Julia Carbajal[b], Bruno Bianchi[a,d], Mariano Sigman[c,e], and Diego E. Shalom[d,e]

[a]Laboratorio de Inteligencia Artificial Aplicada, Departamento de Computación, Universidad de Buenos Aires, Buenos Aires, Argentina; [b]Laboratoire de Sciences Cognitives et Psycholinguistique, Ecole Normale Supérieure, Paris, France; [c]Laboratorio de Neurociencia, Universidad Torcuato Di Tella, Buenos Aires, Argentina; [d]Departamento de Física, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina; [e]CONICET, Argentina

ABSTRACT

When a word is read more than once, reading time generally decreases in the successive occurrences. This Repetition Effect has been used to study word encoding and memory processes in a variety of experimental measures. We studied naturally occurring repetitions of words within normal texts (stories of around 3,000 words). Using linear mixed models to analyze the evolution of fixations over successive repetitions, we observed an interaction between corpus word frequency and repetition. Specifically, we found a decrease in fixation durations in words with low frequency but not with high frequency, and both values converged after five or six repetitions. Furthermore, we showed that repetition of a lemma is not enough to evoke this effect. Our results are in agreement with predictions formulated by the context-dependent representation model, and this adds new arguments to the discussion of the sources of the repetition effect.

Introduction

When a word is read more than once, reading time generally decreases with each successive occurrence. This phenomenon, known as the Repetition Effect, provides a sensitive measure of the content and structure of memory. Results from previous studies on repetition effects shed light on the representation of individual words in memory. Two main explanations were proposed to account for the results: the *abstract view* and the *episodic view* (Bowers, 2000; Raney & Rayner, 1995; Tenpenny, 1995). According to the abstract view, repetition effects result from the priming of word representations independently of the context. Conversely, according to the episodic view, facilitation arises from memory of specific information or events in the text, such as the context or even the physical attributes of the text. However, evidence to support each type of effect seems to depend on experimental tasks and the reader's goal (Carr et al., 1989; Levy & Burns, 1990). As pointed out by Raney (2003), "Given the ample evidence supporting the existence of both abstract and episodic repetition effects, the most appropriate conclusion may be that both abstract-like and episodic-like repetition benefits exist" (p. 19). As a parsimonious integration of both views, Raney (2003) put forward a *context-dependent representation model*.

Raney's proposal builds on van Dijk and Kintsch's model (1983), which postulates three levels of representation: The surface form consists of the wording used in the text and includes lexical and syntactic information of the text, the textbase representation contains the meanings of words, and the situation model integrates the textbase information with prior knowledge. The main assumption of

CONTACT Juan E. Kamienkowski juank@df.uba.ar Departamento de Computación, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Pabellón 1, Ciudad Universitaria, C1428EGA, Ciudad Autónoma de Buenos Aires, Argentina.

Raney's model is that these three levels of representation reflect different degrees of context dependence. Although surface form and textbase representations are context-independent, a well-developed situation model binds the surface form and textbase with general knowledge and inferences to create a unified and context-dependent representation. As a result, the repetition effect within texts also becomes context-dependent. Raney's model states that the size of the repetition effect is determined by properties of both the task and the text. Intermediate size of repetition effects is expected when situation models are not task-relevant or cannot be well formed. Examples of this condition are proofreading texts without comprehension demands (Singer & Halldorson, 1996), and reading passages that are ambiguous unless the title is given with the passage (Levy, 2001). The model predicts a maximum repetition effect when the situation model is task-relevant and complete, and there is an overlap of the surface form, the textbase, and situation models, as in the case of reading for comprehension. On the other hand, a minimum repetition effect is expected when the situation model is task-relevant and complete, but the texts do not refer to the same situations; this is the case, for example, when reading a word a second time, but in a new text with a different topic.

In this work we studied naturally occurring repetitions of words within long texts (stories of around 3,000 words). The richness and variability of this linguistic material allowed us to study the repetition effect in a variety of conditions. In this reading material, some words were repeated many times, which allowed us to study cumulative repetition effects along multiple readings of a word. Raney's model predicts a maximum repetition effect for this case, because surface form and textbase are identical, and the situation model is task-relevant and complete (reading for comprehension). Moreover, all instances occurred in the same situation model, but because our participants read more than one text in one reading session, we were also able to assess the repetition effect of words across different texts. In this case, the context-dependent representation model predicts a very small repetition effect, based on the fact that the surface form and textbase are reinstantiated in a different situation model. Alternatively, the perceptual tuning to repeated visual stimuli was proposed by Vanyukov et al. (2012) as the seed of the frequency effect. This idea is based only on the surface form of the word (i.e., with its physical attributes but not the context). Thus, it is important to note that it does not correspond to an episodic-like representation, which involves both context and physical attributes of the text, or to an abstract-like representation, which is related to the textbase representation. To the best of our knowledge, these predictions have not been tested empirically.

Only a few studies have looked at repeated words within a text (O'Brien et al., 1997; Rayner et al., 1995). O'Brien et al. (1997) compared the gaze duration and probability of fixation on the second reading of a word when it was preceded by another word that was conceptually identical or not. They showed that participants were slightly more reactive when the antecedent target was both identical conceptually and lexically but not when it was just identical lexically. Thus, they only found a repetition effect when the word shared both surface form and textbase but not when the word shared only the surface form. In the present study, we aimed to investigate the complementary situation, that is, the cases in which a word and its repetition shared only the textbase form, and, thus, we analyzed the case of the lemma repetition. We refer to lemma as the canonical form that is associated with a meaning. For example, in English the verb "to play" may be found in inflected forms as "play," "played," "plays," and "playing," all of which correspond to the lemma "play," which is the only form found in the dictionary. In the case when lemma is repeated, Raney's model predicts no effect (or at least a smaller effect than the one predicted for same-word repetition), because there is no overlap between instances in the surface form. O'Brien et al. (1997) also explored the effect of distance between successive instances of a word and found reactivation after 10 to 60 words. This result is also accounted for by Raney's model: As long as the three levels of representation are identical, the repetition effect is expected to occur. Hence, no distance effect is expected. Our goal is to extend this observation to a wider range of distances; in our study, the amount of words between the first and second appearance of a word ranged between 1 and 100 words.

Rayner et al. (1995) presented results on the interaction between repetition and frequency in the first four repetitions of a word within the same text. Their study was a reanalysis of the data from

Raney and Rayner (1995); the authors found a decrease for both first fixation and gaze duration with repetition, and this effect seemed to be larger for low- than for high-frequency words. Several experimental paradigms have explored whether frequency and word-repetition effects are additive or interactive. In this context, the effect of two variables is referred to as interactive when their simultaneous influence on a third (fixation durations) cannot be explained by the simple addition of the pure effects of the two variables. Conversely, two effects are additive when the interaction between them is zero. According to the additive factors logic (Sternberg, 1969), additive variables (as opposed to interactive variables) are presumed to influence different stages of processing. On the one hand, Chamberland et al. (2013) argue in favor of additivity (no interaction) in paradigms in which the same paragraph is read twice in succession. In a similar paradigm, Raney and Rayner (1995) found no interaction but discussed a floor effect as a potential explanation for this lack of interaction. In this context, a floor effect appears when a minimum time is required to process the simplest (short, high frequency) words. Thus, the fixation duration remains constant and is not affected by further facilitation of the word. On the other hand, Rayner et al. (1995) examined fixation times for multiple occurrences of low- and high-frequency words within a single paragraph, in a reanalysis of the same data. Their results suggested that the high-frequency words reached a floor before low-frequency words. Similarly, many word-list studies showed an interaction between frequency and repetition. In particular, Kinoshita (2006) tried to solve this inconsistency and showed that both additive and interactive effects were present: Their relative magnitude depended on the familiarity of the subjects with very low-frequency words. We aimed to further dissect the interactive/additive effects of frequency and word repetition. The use of long texts ($>$1,000 words) presents an ideal condition for studying this issue, due to the wide range of frequencies and appearances of words that are naturally repeated.

Other lines of research have studied the effect of predictability on eye movements. Predictability is usually defined as the probability that the next word in a sentence is guessed, given only the prior words of the sentence (i.e., incremental cloze task procedure; Taylor, 1953). A robust finding was that fixation durations decreased with increasing predictability and that word-skipping probability increased with increasing predictability (Balota et al., 1985; Ehrlich and Rayner, 1981; Rayner and Well, 1996). When a word is repeated within a context, its predictability is expected to rise, primed in part by the previous appearances and by the context. However, not having empirical data of predictability for long texts, we can only speculate. We may consider the repetition effect as a result of increased predictability. Predictions derived from this hypothesis are similar to those derived from the context-dependent representation model. Raney's model predicts a maximum repetition effect when the situation model is complete. This corresponds to a strong context that would lead to high predictability values. A minimum repetition effect is expected when texts do not refer to the same situations, where predictability values would also be low.

The present study aimed to deepen the understanding of repetition effects by using a material that was quite different from both word lists and sentence and paragraph repetition. The use of everyday long texts also strengthens the generality of repetition effects. Based on considerations of the context-dependent representation model, a series of predictions can be extracted. Repetitions within a thematic context (in our work instantiated in each individual text) are expected to have a stronger effect than repetitions that occur across different contexts. The repetition of the specific word form is expected to show a stronger effect than the repetition of the canonical form (i.e., lemma repetition). The distance between successive instances of a word is expected to have no or a negligible effect on eye movements. Finally, a higher cumulative repetition effect is expected for low- than for high-frequency words. Overall, we seek to dissect further the interactive/additive contributions of frequency and repetition. Therefore, we evaluated the extension of the benchmark list of covariates by introducing word repetition as a reliable and strong predictor of fixation durations.

Methods

Subjects

Thirty-six healthy subjects (11 women; age range 20 – 40 years; $M = 24.9$, $SD = 3.8$) participated in a 2-hour reading experiment. All participants were native Spanish speakers and had normal or corrected-to-normal vision. All participants were compensated with 30 Argentine pesos for 2 hours of participation.

Procedure

Short stories were presented on a PC monitor (black Courier New bold font; 0.44° letter width; 0.31° minimum letter height; gray background). Each story was presented throughout several screens, with 10 lines of text at a time (double-spaced, 1.6° interline spacing; 55 character maximum per line), plus an extra screen with the title and author presented at the beginning of the session. Subjects were instructed to read at their own rate, moving forward or backward in the screen sequence by pressing the right and left arrow keys, respectively. No instructions were given to suppress eye blinks. Each participant completed sessions of 2 hours (3 – 4 texts, $M = 3.4$). Texts were assigned pseudo-randomly to participants to achieve a similar number of readings of each text. The average reading rate was 202.7 words per minute ($SD = 32.3$), which is within the normal reading rate (Legge et al., 1985; Pelli et al., 2007; Rayner, 1998). Reading for comprehension was stated in the instructions to the participants and reinforced by telling them that comprehension questions were to be made after each text. Subjects answered five questions regarding the contents of each text, which was used to determine comprehension level. We obtained an average of 4.7 correct answers and a minimum of 3.

Equipment

Participants were seated in front of a 19-inch screen (SyncMaster 997 MB, 1,024 × 768 pixels resolution, 100 Hz refresh rate; Samsung, Suwon, Korea) at a viewing distance of 65 cm, subtending an angle of 29.3 degrees horizontally and 22.5 degrees vertically. A chin rest that was aligned with the center of the screen prevented head movements. An EyeLink 1000 eye-tracker (SR Research Ltd., Ottawa, Ontario, Canada) was used to record gaze locations of both eyes during reading at a sampling rate of 1 kHz. Nominal average accuracy is 0.5 degrees, and space resolution is 0.01 degrees root mean square, as given by the manufacturer. The participant's gaze was calibrated with a standard 13-point grid for both eyes. Two nine-point validations were run before and after each text. Based on these validations, the best calibrated eye was selected for each participant. The averaged accuracies for their best eye before and after reading the texts were 0.42 ± 0.20 degrees and 0.63 ± 0.24 degrees (mean ± SD), respectively. All eye movements were labeled as fixations, saccades, and blinks by the eye-tracker software using the default thresholds for cognitive experiments (30°/s for velocity, 8,000°/s^2 for acceleration, and 0.1° for motion) (Cornelissen et al., 2002). Presentation of stimuli was developed using Matlab (http://www.mathworks.com/) and Psychophysics Toolbox Version 3 (Brainard, 1997).

Buenos aires corpus

A corpus of 10 short texts (mean length 3,312 words, min = 1,975, max = 4,640; mean number of screens 39, min = 24, max = 59) written in Spanish was selected specially for the experiment. The whole corpus contains a total of 33,120 words. Texts were checked to have no typographical errors, and all words were classified by grammatical class and lemma. These word properties were taken from the Spanish LexEsp corpus (Sebastián-Gallés et al., 1998) and curated manually. To verify the properties of this corpus, an automated grammatical categorization was run using Freeling (Padró & Stanilovsky, 2012). For content words (nouns, verbs, and adjectives), we found an agreement of 94% between both classifications. Content words represented 46.5% of the words (of those, 46% were nouns, 38% were verbs, and 15% were adjectives). The texts contained an average of 126 ($SD = 43$) sentences, and mean sentence length was 28.2 ($SD = 9.5$) words. Each story contained an average of 955 ($SD = 170$) unique content words. From those unique words, an average of 718 ($SD = 122$) appeared only once, 237 ($SD = 67$) appeared at least twice, 108 ($SD = 40$) appeared at least three times, 61 ($SD = 27$) appeared at least four times, and 40 ($SD = 20$) at least five times.

Word length ranged between 1 and 19 letters ($M = 4.6$, $SD = 2.8$). Printed frequency was taken from the Spanish LexEsp corpus. Word frequency ranged from 0 to 46,567 per million; the mean log frequency (incremented by 1) was 2.6 ($SD = 1.5$). Particularly, length of content words ranged between 1 and 17 letters ($M = 6.6$, $SD = 2.4$), and word frequency ranged from 0 to 6,003 per million; the mean log frequency (incremented by 1) was 1.4 ($SD = 1.0$). Further details are presented in the Supplemental Table 1. Supplemental material is available at http://reading.liaa.dc.uba.ar/.

Data set

Eye movement data from 105 readings among all participants (we defined a reading as one text read by one subject) was screened for blinks and track losses. Fixations shorter than 50 ms and longer than 1,000 ms were removed from the analysis. After this screening process, fixations were assigned to their respective word/line. The procedure involved adjusting the limit between each pair of lines of texts visually to assign each fixation to the correct line. We felt compelled to perform this adjustment because a few fixations were located near the limit in-between lines; a single miss-assigned fixation in the vertical line can change the assignment of a whole line of text as first-pass reading. The overall mean displacement of each line limit was 0.25 degrees, about half the height of a character, and about 15% of the interline spacing. An average of 57 fixations was reassigned in each reading, which corresponded to 1.4% of all fixations.

We finished with an original dataset that contained a total of 351,680 presented words. From these words, 163,713 were content words (nouns, verbs, and adjectives). Our eye-movement dataset

Table 1. Baseline and repetition models for gaze durations.

	Gaze Duration								
	Baseline Model			Simple Repetition Model			Full Repetition Model		
Random Effects									
Groups	Variance	SD		Variance	SD		Variance	SD	
Word ($n = 5,259$)	0.003	0.052		0.003	0.051		0.003	0.050	
Participant (n = 36)	0.002	0.045		0.002	0.045		0.002	0.045	
Text (n = 10)	< 0.001	0.013		< 0.001	0.012		< 0.001	0.013	
Residual (n = 83,903)	0.030	0.174		0.030	0.174		0.030	0.174	
Fixed Effects									
Covariate	*M*	*SE*	*t*-value	*M*	*SE*	*t*-value	*M*	*SE*	*t*-value
(Intercept)	2.389	0.009	277.1	2.387	0.009	277.8	2.385	0.009	277.0
Launch site	0.037	0.001	57.2	0.037	0.001	57.3	0.037	0.001	57.4
1/Length (N)	− 0.628	0.020	− 31.3	− 0.619	0.020	− 31.0	− 0.624	0.020	− 31.6
1/Length ($N - 1$)	0.076	0.005	16.4	0.075	0.005	16.4	0.075	0.005	16.3
Freq (N)	− 0.029	0.001	− 21.3	− 0.028	0.001	− 20.2	− 0.025	0.001	− 17.7
Freq ($N - 1$)	− 0.015	0.001	− 20.0	− 0.015	0.001	− 19.9	− 0.015	0.001	− 19.7
Freq (N) × 1/Length (N)	0.374	0.018	20.6	0.380	0.018	21.1	0.346	0.018	19.1
Positional Effects									
rpl	0.028	0.004	7.7	0.028	0.004	7.7	0.028	0.004	7.8
rpt	− 0.107	0.024	− 4.4	− 0.086	0.025	− 3.5	− 0.095	0.025	− 3.9
rpt^2	0.191	0.057	3.4	0.158	0.057	2.8	0.171	0.057	3.0
rpt^3	− 0.103	0.038	− 2.7	− 0.085	0.038	− 2.3	− 0.091	0.038	− 2.4
rps	− 0.014	0.003	− 5.6	− 0.014	0.003	− 5.5	− 0.015	0.003	− 5.8
Sentence border	0.029	0.007	4.3	0.030	0.007	4.5	0.030	0.007	4.6
Repetition Effects									
1/NREP (N)				0.017	0.003	6.4	0.028	0.003	9.6
1/NREP (N) × Freq (N)							− 0.029	0.003	− 10.1

contained a total of 414,899 fixations. Because some words were skipped, the dataset contained 147,467 content words fixated at least once. Some of these words were fixated for the first time *after* a fixation on a subsequent word (i.e., the reader went back and read the word that had been skipped). This left us with 136,235 content words that were fixated in progressive, first-pass reading. We further excluded from this dataset: first and last word of each text line (and paragraph) (101,596 remaining words) and fixated words whose incoming and/or outgoing saccade were not contained in the same line. Thus, the final dataset contained a total of 83,903 fixated words. The repetition number was computed as a property of each word in the original texts. Hence, the repetition number was not affected by this elimination procedure. For each fixated word in progressive, first-pass reading, the *Gaze Duration* was defined as the sum of all subsequent fixation durations on the word before any other word was fixated. These measures were computed using the *em2* package (Logacev & Vasishth, 2013) for R language for statistical computing (version 3.0.2; R Core Team, 2013).

Linear mixed models

In this project the *lmer* function included in the *lme4* package (version 1.1-8) (Bates et al., 2013) was used for estimating fixed and random coefficients. This package is available for R language for statistical computing (version 3.0.2; R Core Team, 2013). Maximum likelihood statistics were used for model comparison with different fixed effects and identical random effects; restricted maximum likelihood statistics were used for an estimation of fixed and random effects in the final model. For assessment of (differences in) goodness of fit, the lmer program provides the Akaike Information Criterion ($AIC = -2\ logLik + 2nparam$; decreases with goodness of fit) (Sakamoto et al., 1986), the Bayesian Information Criterion ($BIC = -2\ logLik + nparam\ log(Nobs)$; decreases with goodness of fit) (Schwarz, 1978), the log likelihood (*logLik*; increases with goodness of fit), and, in the case of model comparison, the likelihood ratio. The AIC and BIC values correct the log-likelihood statistic for the number of estimated parameters and the number of observations to avoid overfitting during the process of model selection.

The critical factor in this study was the number of repetitions of the word. Special attention was paid to its interaction with frequency, because this has been described in other paradigms that compared the reading of consecutive paragraphs (Hyönä & Niemi, 1990; Raney & Rayner, 1995; Raney et al., 2000) or that use sequential foveated presentations (Van Petten et al., 1991).

The linear mixed models (LMMs) included a number of other covariates known to affect fixation durations. Launch site is known to have strong effects on duration of fixations (Heller & Müller, 1983; Pollatsek et al., 1986). We included inverse length and log frequency of word *N*-1 and *N* as covariates to facilitate the comparison with models previously reported (Fernández et al., 2014; Kliegl et al., 2006). A traditional variable used in coding predictability in ERP research has been the ordinal position of a word in the sentence (Dambacher et al., 2006; Van Petten et al., 1991), which was also reported to be a significant covariate in LMM analysis of eye movements (Fernández et al., 2014; Kuperman et al., 2010; Pynte et al., 2008a, 2008b). Furthermore, the first and last word in the sentence were included as a categorical variable. Because the sentence did not necessarily match a line, we added the position within the line as a separate covariate. Finally, because these are large texts in comparison with the isolated sentences used in other experiments, we introduced ordinal position of a word in the text as another covariate. The positions within the sentence, the line, and the text were rescaled to the [0 1] interval and named as relative position in the sentence (*rps*), relative position in the line (*rpl*), and relative position in the text (*rpt*). All covariates were centered so that the intercept estimated the mean log duration.

For the LMMs, regression coefficients (bs), standard errors (SEs), and t-values ($t = b/SE$) are reported. There is no clear definition of "degree of freedom" for LMMs, and therefore precise *p* values cannot be estimated. However, in general, given the large number of observations, subjects, sentences, and words considered in our analysis and the comparatively small number of fixed and random effects estimated, the t-distribution is equivalent to a normal distribution for all practical purposes (i.e., the contribution of the degrees of freedom to the test statistic was negligible). Our criterion for referring to an effect as significant was $t = b/SE > 2.0$. The significance on fixed effects was checked with the *confint()* function implemented in the lme4 package. This function estimates confidence intervals by computing a likelihood profile and finding the appropriate cutoffs based on the likelihood ratio test. In all 183 tests both statistics led to the same decision.

We specified LMMs such that they yielded estimates for variance components that were associated with intercepts for participants, texts, and words. The models account for dependencies between fixations due to the clustering that was associated with these three partially crossed, random factors. We also studied the influence of distance between repeated words, the congruences between repetition of lemma and repetition of form and between the repetition within the text or within the entire experiment (across the 3 or 4 texts read by the participant). Unless otherwise noted, the number of samples considered for each model was 83,903.

Results and discussion

Our analysis focused on the effect of repetition of single words within long texts. With that purpose in mind, we analyzed first-pass reading of content words. Because there were only a few studies that used LMMs of eye movement in reading either long or Spanish texts, we began by building a baseline model, which included that are commonly known: the (log) frequency and (inverse) length of the fixated word (N) and the preceding word ($N - 1$), and the launch site. We evaluated the addition of some new positional effects, such as the position of the word in the line, the sentence, and the text. We also added participant, text, and word as random effects. It was built progressively by adding one covariate at a time and keeping only significant effects (see Table 1, first column, and Supplemental Table 2 for further details). These models were based on the final dataset that contained 83,903 fixated words (see Methods).

Repetition model

The baseline model did not differentiate whether words were presented for the first time or were repeated one or many times in the texts. To include this property in the models, we defined a new variable, the repetition number NREP, and assigned a value to each individual word: 1 for the first appearance of the word within each text, 2 for the second, and so on. It is important to mention that NREP accounted for repetitions within each individual text. Cumulative repetition across texts read by each participant is considered in the following paragraph. This newly introduced repetition variable had a considerable impact on gaze durations (Figure 1), and this effect saturated after about five to six repetitions. Thus, as a first step, we compared the effect of NREP with the inverse transformation

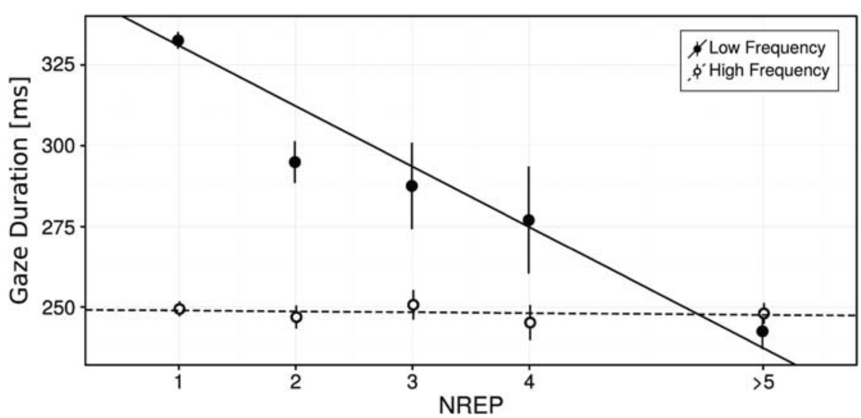

Figure 1. Interaction of repetition number with frequency. Gaze as function of NREP, for high (>66th percentile) and low frequency (<33rd percentile).

(1/NREP) to account for this saturation. The inclusion of either NREP or 1/NREP as a new covariate improved the baseline model, but this last model was significantly better than the model with a linear NREP, which supports the idea that the repetition effect saturates after a certain number of repetitions ([AIC BIC]: from linear: [− 51,548, − 51,380] to inverse: [− 51,578, − 51,410]). Therefore, we used the inverse transformed for the rest of our work. The addition of this new covariate to the Baseline Model (Table 1, second column) improved the goodness of fit significantly, as assessed by the likelihood ratio statistic ($\chi^2(1\ \text{df}) = 40.2$, $p < 5 \times 10^{-10}$) and a decrease of both the AIC and BIC indices ([AIC BIC]: from [− 51,540, − 51,381] to [− 51,578, − 51,410]). The 1/NREP had a significant positive effect on gaze durations ($t = 6.4$), and the direction of the 1/NREP effect was consistent with shorter duration of successive appearances. All other covariates remained significant in the repetition model (Table 1, second column).

Because participants read between three and four texts (which corresponded to different contexts), we were able to assess word repetition both "within a text" (i.e., NREP) and "across texts." The former only accounts for repetitions in the same context, whereas the latter also accounts for repetitions across completely different contexts. The *context-dependent representation model* proposed by (Raney, 2003) states that the repetition effect is maximal when the situation model is task-relevant and complete, and there is a strong overlap between the context of the repeated words, which is the case for repetitions within the same text. However, because there is no overlap between contexts of repeated words across different texts, a very small repetition effect is expected for the repetition "across texts." On the other hand, the perceptual tuning of repeated visual stimuli hypothesis, as proposed by Vanyukov et al. (2012), suggests a context-independent repetition effect. Thus, we compared the goodness-of-fit when we include the repetition "within a text" with the repetition "across texts" to the baseline model. Accordingly, because both types of repetitions were identical within the first text, we decided to exclude them from this analysis. This created a dataset of 53442 fixated words. Both AIC and BIC were smaller for "within a text" than "across texts" ([AIC BIC]: "within a text": [− 51,578, − 51,410], "across texts": [− 51,562, − 51,394]), which indicates that repetitions that occurred within the same context contribute to a better model. This suggests that the repetition effect was influenced strongly by the context, and it cannot be attributed only to a perceptual effect (i.e., just the number of repetitions of the surface form of the word).

Previous studies showed that priming effects can be observed not only for repetitions of the same word but also when the root, or words that are morphologically or semantically similar, are presented as a prime (Marslen-Wilson et al., 1994; Neely, 1977). We address whether the repetition effect requires strict word repetition or if it can be achieved by different words that share the same root (*repetition of lemma*). To this aim, we compared the repetition effect on gaze durations of successive appearances of lemmas in different words (lemma repetition effect) with the effect on words that were repeated exactly (word repetition effect). The repetition of lemma effect was significant ($t = -6.7$), and the goodness of fit was not statistically different from the word repetition model ([AIC BIC]: "words": [− 51,578, − 51,410], "lemmas": [− 51,583, − 51,415]). However, this analysis has a potential limitation. The word repetition is included in the lemma repetition, because repeated words usually share their lemma. For this reason, we reduced the dataset with the aim of avoiding differences in the history of a given repetition, that is, if previous appearances were repetitions of the lemma, a word or a mixture of both. Thus, we kept only the first and second appearance of repeated words and lemmas, which resulted in a dataset that contained 53,589 fixated words. We assigned a dummy variable to each word, with one of three levels: LW1 referred to the first appearance of a word, W2 referred to the second appearance and repetition of the word form, and L2 referred to the second appearance and repetition of the lemma but not the word form. For example, if the words "tree" and "tree" appeared in this order, they would be labeled LW1, W2. Subsequent repetitions of either the lemma or the word would be discarded. Similarly, if the words "plant" and "plants" appeared in this order, they would be labeled LW1, L2. And subsequent repetitions would be discarded from this dataset. We added this factor to the same baseline model in a treatment design, with LW1 as a baseline. Interestingly, we observed an effect of repetition between words (W2 − LW1: $t = -5.2$) but not between lemmas (L2 − LW1: $t = -1.5$). Note that in this case, the sign of the word repetition effect is inverted, because we considered a linear-like effect of repetition instead of the inverse of the repetition number. The same general effects of Table 1 were present in this reduced dataset.

Based on the perceptual tuning hypothesis to specific word forms (Vanyukov et al., 2012), the memory for a given word should fade away after a certain amount of time, and the repetition effect should decay with the time elapsed between repetitions. On the other hand, the context dependent representation model (Raney, 2003) specifies the requirement of a complete situation model for the repetition effect to be maximal. In many examples this completeness is achieved with few sentences or even only a title (Levy, 2001). If this was true in the present study, the situation model must have been well-developed within each text, and repetition effects would be expected to be maximal, independent of the time elapsed between repetitions. In this scenario we aimed to assess if there is a timescale in which the repetition effect fades out, that is, if the repetition effect can occur even after a long interval. As a simple estimate of the time elapsed between successive readings of a word, we measured the distance (in number of intermediate word) to the previous appearance of the same word. The values of the distance ranged between 1 and 4,241 words ($M = 673$, $SD = 752$). We excluded the first appearance of a word, because the distance to previous repetition is not defined, resulting in a total of 30,101 fixated words in the dataset. All the previous effects were preserved in this reduced dataset. We failed to find a significant effects of distance or of an interaction between NREP and distance (distance: $t = 0.5$, 1/NREP × distance: $t = 0.2$). However, this analysis has a potential limitation when a word is repeated more than twice. The effect not only depended on the distance to the preceding appearance but also depended on the history of all previous repetitions of that word. Thus, we repeated this analysis, but only for the second repetition (NREP = 2), which resulted in 12,638 fixated words, and we also failed to find any significant effect of distance ($t = 1.0$). Finally, inclusion of the distance as a predictor assumes that the effect is monotonically increasing or decreasing. Because this might not have been true, we also explored adding a quadratic term, which was also not significant ($t = 0.0$). These negative results suggest that the repetition effect is independent of the distance, at least in the characteristic timescale of these texts, where words were repeated typically at 100 to 1,000 words.

Interactions with frequency

One of the landmarks of the repetition effect is its interaction with the frequency of the repeated word. The following step was to determine if this reduction in gaze duration also implied a change, not only in the duration of a gaze to a word but also on how other factors determined the time spent on a word. Thus, we explored the interaction between the effects of frequency and word repetition. The dependence of gaze duration on repetition clearly showed different patterns for the two extreme terciles of frequency (Figure 1). High-frequency words were unaffected by repetition, but gaze duration to low frequency words decreased with repetition. Furthermore, low-frequency words reached the values of high-frequency words after five or six repetitions.

To quantify this observation, we expanded the model by adding the interaction of 1/NREP with the frequency of the present word (Table 1, third column). The addition of this interaction to the Repetition Model improved the goodness of fit significantly, as assessed by the likelihood ratio statistic (χ^2(1 df) = 100.8, $p < 5 \times 10^{-16}$) and a decrease of both the AIC and BIC indices ([AIC BIC]: from [−51,578, −51,410] to [−51,676, −51,499]). The significance and direction of the other covariates remained the same, including the 1/NREP ($t = 9.6$), but the new interaction ($t = -10.1$) had a negative estimate. This indicates that less frequent words have larger repetition effects (Figure 1).

In this context, it is important to determine whether this high- versus low-frequency transition is gradual or abrupt; therefore, we partitioned the words into 10 frequency deciles. We performed simple linear regressions of gaze duration as a function of 1/NREP for each partition (Figure 2, a–c). The effect of repetition decreased gradually with increased frequency, as shown by a smooth change in the slope as frequency increased (Figure 2b). The intercept corresponds to the asymptotic duration after many repetitions, because 1/NREP tends to zero for large NREP values. An intercept that is nearly constant

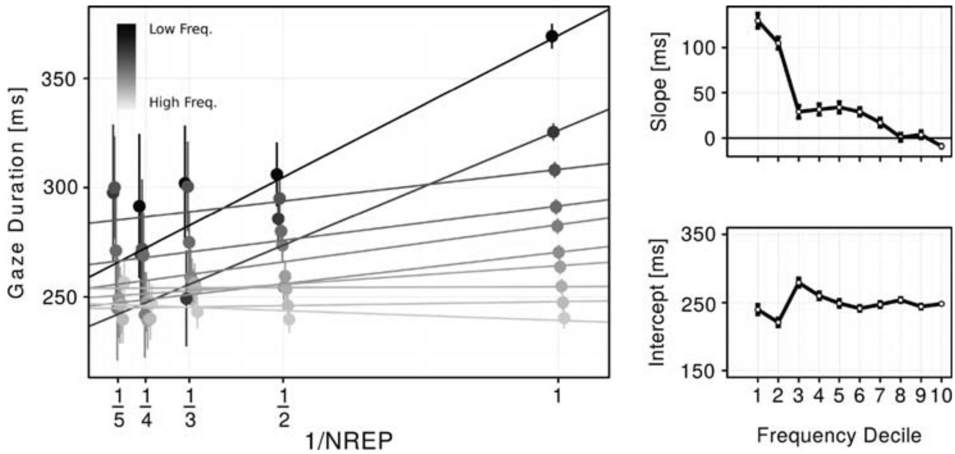

Figure 2. Interaction of repetition number with frequency deciles. (a) Gaze as function of 1/NREP, as it is used in the linear mixed models, separated in 10 frequency deciles, dark lines correspond to low frequency deciles (median frequency of each decile was 0.8, 2.6, 6.6, 12.5, 22.5, 40.2, 67.3, 113.9, 254.9, and 830.7 occurrences per million). (b) Slopes (obtained from linear fits of Gaze durations by 1/NREP) as function of decile frequency. (c) Intercepts (obtained from linear fits of Gaze durations by 1/NREP) as function of decile frequency.

indicates that the asymptotic gaze duration was nearly independent of the frequency after many repetitions (Figure 2c). This asymptotic value corresponds precisely to the gaze duration of high frequency words.

To quantify this observation, we analyzed a model adding the interaction of 1/NREP with the decile number (the same partition of Figure 2, a–c). Consistent with Figure 1, we found that the lowest frequency deciles exhibited a significant effect of 1/NREP and that this effect decreased progressively in value and, then, completely disappeared in the seventh decile and beyond (Supplemental Table 3).

Although these results indicate a clear interactive effect between repetition and frequency, there are still two possible interpretations: Either the effect is truly interactive in nature, that is, that repetition affects how frequency of words is processed. Or, the effect is additive but an interaction arises from a floor effect; that is, there is no benefit of repetition for high frequency words because fixations cannot last less than a certain minimum value.

As a first approach, if there was an additive effect masked by a floor effect, the curves for each decile in Figure 2a should have been parallel until each one reached the floor value. On the other hand, if there was an interactive effect, it should have decreased with different slopes and it should have been more pronounced for lower frequency words. Thus, we repeated the analysis but only for the first and second appearances of each word and only for low frequency words (median-split) to avoid the floor-effect region (a total of 9,679 fixated words were included in this analysis). The interaction was still significant in this reduced dataset ($t = -2.6$), which indicates an interaction even in the region where the floor effect is not present.

We showed that repetition of high-frequency words did not lead to shorter fixations, probably because saccade programming takes a certain minimum amount of time. However, it might still be possible to see benefits of repetition on the following word, if repeated high-frequency words produced more parafoveal processing and, hence, reduced fixations on word *N*. In our analytic framework, this is equivalent to evaluating the effects of word *N*-1 on the gaze durations on word *N*. In particular, we wanted to evaluate the addition of an interaction between 1/NREP(*N*-1) and FREQ(*N*-1) to the baseline model. We observed a significant effect of the interaction between 1/NREP(*N*-1) and FREQ(*N*-1) on word *N* ($t = -8.8$). This result held separately for both function ($t = -3.1$, 56,607 fixated words in the dataset) or content ($t = -2.4$, 27,296 fixated words) words in the position *N*-1 and also for low-frequency words (median-split) that were not affected by the floor effect of the word *N* ($t = -6.1$,

$n = 41{,}887$). Finally, we evaluated the effect of 1/NREP(N-1) for high frequency N-1 words (fifth quintile) ($n = 5{,}469$), and found no effect of 1/NREP(N-1) ($t = -0.5$).

These partial results taken together suggest that repetition not only reduced fixation durations, but it also changed interactively changes how words were processed. These results add statistical support and deepen previous observations on word repetition within the same text (Rayner et al., 1995).

Other eye movement measures

Studies on reading have relied on different measures of fixation durations: First Fixation Duration, Single Fixation Duration, Gaze Durations, Word Skipping, and other measures, which include regressions and second-pass reading (Rayner, 1998, 2009). Each of those measures highlight some aspect of word processing, but they are not necessarily independent. Thus, we also investigated the robustness and generality of the main effects by repeating the analysis with other measures (Supplemental Table 4). Because multiple dependent variables were studied, we applied the Bonferroni correction for multiple comparisons. Thus, the threshold of significance was lowered, dividing the standard value of 0.05 by the number of different eye measures (von der Malsburg & Angele, 2016) (or six in the present case). In the case of LMMs, this entails changing the t-value significance threshold from 2.0 (corresponding to an uncorrected threshold of 0.05 in a Student's t-distribution with infinite degrees of freedom) to 2.7 (corresponding to a Bonferroni corrected threshold of $0.05/6 = 0.008$).

Our results showed that gaze durations decreased when a word was repeated. However, it is not clear what changed exactly in the reading behavior. Gaze duration can be decomposed into the duration of the fixations and the probability of refixation. Hence, increased gaze durations could be due to the increase in either of these two variables. We tested both hypotheses. First Fixation Duration showed significant effects of both NREP ($t = 7.8$) and a significant interaction between NREP and frequency ($t = -5.4$).[1] In addition, we measured the Number of Fixations in First Pass, that is, the number of fixations that composed the gaze. We also found a significant effect of NREP ($p < .001$) and a significant interaction between NREP and frequency ($p < .001$) on Number of Fixations in First Pass. These results indicate that not only are the fixations shorter when a word is repeated, but the word is also refixated less often.

One might also speculate that if the processing load was reduced for repeated words, in some cases, these words could be preprocessed parafoveally in the fixation to the previous word and consequently skipped. We built a model with the probability of a skip in first-pass reading as the dependent variable.[2] For this model, we also included the content words that were not fixated in first-pass reading but still excluded first and last word of each line ($n = 89{,}288$). We found a nonsignificant (Bonferroni corrected) effect of NREP ($p = .05$) and a significant interaction between NREP and frequency ($p < .005$) on the probability of skip.

Finally, we explored two measures of late processing, Regression Path Duration and Rereading Time (Clifton et al., 2007; Demberg & Keller, 2008; Rayner, 1998, 2009). Regression Path Duration or Go-Past Duration is the sum of all first-pass fixation durations on all preceding positions (including the present word), from the first fixation on the present word until the first fixation on anything to the right of the present word. Regression Path Duration showed a significant effect of NREP ($t = 9.2$) and a significant interaction between NREP and frequency ($t = -10.6$). Rereading Time is the sum of all second-pass fixation durations. It excludes the first-pass reading time, which is included in the Regression Path Duration. Thus, we excluded the words that were fixated only in the first pass

[1]As expected from results on Gaze and First-Fixation Durations, Single-Fixation Duration also showed significant effects of both NREP ($t = 8.5$) and the interaction between NREP and frequency ($t = -5.8$).

[2]Because the probability of skip, as defined in our dataset has only two possible values, that is, 0 if the participant fixated on a given word in that text, and 1 if the participant did not, we used a Generalized LMM with kernel = binomial, instead of the LMM that was used for continuous dependent variables.

($n = 20{,}391$). The repetition effect ($t = 3.3$) and its interaction with frequency ($t = -3.7$) were also significant when considering Rereading Time.

Hence, the processing of a word is affected at many levels by its previous appearances in the same text, as indicated by the different eye measures considered here. The effect of repetition has been observed in many different eye measures, and it remains significant after treatment for multiple comparisons (von der Malsburg & Angele, 2016). Also, the significant interaction observed in various measures further supports the previous conclusion that the effects of frequency and repetition were interactive in nature, beyond the floor effect. The lack of repetition effect on skips, in addition to the significant effect on late measures, suggests that repetition affects both lexical processing and integration processes, because it does not affect the rate of skipping, although it shortened all the processes that occurred after the first fixation.

Conclusions

We investigated how fixation durations changed when words were repeated naturally in long texts. The main novelty of our work is that our subjects read long stories, which allowed us to investigate how context-independent, word parameters (length and frequency) interacted with local variables of a natural text (whether the word or similar words had been repeated previously, the relative position in the sentence, in the line, and in the text). Our main findings are as follows: (1) the repetition effect shortened the fixation durations; (2) this effect saturated after five or six repetitions; (3) the effect of repetition interacted with the effect of frequency; (4) this interaction between frequency and repetition did not take place due to a floor effect; (5) the effect of repetition *within* a text was stronger than *across* different texts; (6) the repetition of lemmas was not enough to evoke this effect; and (7) the distance between repetitions did not have a significant effect on fixation durations.

The first important conclusion of our study is that the dependence of word fixation duration with canonical factors (frequency) changed during fluent reading in a way that could be accounted for by repetition. Furthermore, this dependence was observed in essentially all eye movement measures. It implies that the effect was not only a speeding up of the process or an additive gain of some hundredths of a seconds but, instead, that the mechanisms of reading themselves (memory storage and retrieval, word-recognition) were affected by the semantic context and word repetition effects. More specifically, the strong interaction between frequency and repetition indicates that low and high frequency words were affected differently by repetitions. Our results show that the highest-frequency words were almost insensitive to repetition. In contrast, fixation durations on low-frequency words were shortened as repetitions proceeded and converged asymptotically to the durations of high-frequency words. These results are in line with Rayner et al. (1995), but contrast with Raney and Rayner (1995) and Chamberland et al. (2013), who found no interaction between frequency and repetition in two successive readings of the same paragraph.

One possible explanation for these divergent results could be attributed to differences in the paradigms used, in particular, differences in predictability. When a text is read twice, every single word is repeated, and thus the predictability of the target words (and all the others) is exceptionally increased. Even more drastic is the effect of repeating the paragraph but changing only the target word, because this change may create a mismatch that might give rise to processes other than pure repetition. In our study, participants read normal texts and, thus, predictability values rose naturally. On the other hand, in those studies, the datasets consisted of a few different items (14 or 16) per word frequency condition, with two repetitions and two word frequency conditions. In contrast, our paradigm included the complete set of repeated content words of the texts, which allowed us to produce a larger and more natural dataset and to explore wider ranges of frequencies and repetitions. Smaller ranges of repetitions might not have been enough to detect the interaction between frequency and repetition, and it would certainly not have been enough to see the saturation. Moreover, it was suggested that the frequency range of the high-frequency words may have been responsible for the interaction (Bodner & Masson, 2001; Kinoshita, 2006). When high-frequency words were chosen to be in the range of 40 to 60 occurrences per million, no interaction was found, but when high frequency words were in the range of 200 per million, interactive effects became significant (Kinoshita, 2006). The lack of significant interactions between frequency and repetition found in these investigations could be due to limitations in the frequency range and in the range of repetitions. In this regard, another possible explanation could be that the interaction between frequency and repetition was a direct consequence of the floor effect. There is a minimum fixation duration required to process even the shortest, repeated, and highest-frequency words, because saccade planning takes certain amount of time (Findlay & Walker, 1999). The dependence on repetition of high and low frequency is hence different, which results in a significant interaction between frequency and repetition. Our results indicate that the interaction was preserved even in the region far from the floor-effect values (low frequency and less than two repetitions). These results, together with the influence of the frequency and repetition of the previous word, suggest that repetition of a word changed interactively how the words of various frequencies were processed.

One important take-home message of this work is that repetition must be certainly taken into account when designing experiments that rely on low frequency words, which are used repeatedly within an experiment. After repeated appearances in a single thematic context, low-frequency words may become indistinguishable from high frequency words, at least with respect to their gaze duration patterns.

Our results agree in general with the predictions formulated by the context-dependent representation model (Raney, 2003). The largest effect of repetition is expected to be found for repetition within a text, when the situation model is task-relevant and complete and when there is an overlap between the contexts of the repeated words. On the other hand, in repetitions across texts there is no overlap between contexts of repeated words and, hence, a smaller repetition effect is expected. This is the pattern of results that we found. When the lemma is repeated, although the textbase is repeated, the surface form is different. Therefore, a smaller effect is expected, as compared with word repetition. This is the case in our results. We also found no distance effect on the duration of fixations, after using a wide range of distances (1 – 1,000 words). This is also supported by the model, because all three levels of representation are identical, a repetition effect is expected to occur independently of the distance between words.

A second aspect that we can only begin to understand is the mechanism by which words that are repeated are read faster. One hypothesis that derives from the observed word specificity of the repetition effect is that a word form remains in memory (not necessarily explicit working memory) for a typical amount of time. If the same word is presented later, the mnemonic effect results in faster reading of such word. Our results show that within the time window explored here (100 – 1,000 words, or about 1 – 10 minutes), the size of the temporal interval between two repeated words did not affect the change in fixation time. Similarly, O'Brien et al. (1997) found no effect of distance on eye movements between first and second appearance of a word (10 – 60 words). A possible interpretation of this result is that memory leak happens on a time scale longer than 10 minutes. This is possible, but it would point to a rather persistent memory for a specific word form. That is, it would be quite surprising that if the word *garden* is presented in the text, 10 minutes later, a repetition of that word is read faster than in a first presentation, but the word *gardening* is not.

Another possibility, which our data seem to refute is a pure semantic priming explanation (Neely, 1977). Semantic priming has been observed in several experiments, which suggests that a word primes an entire set of words that have some degree of semantic similarity. Within the context-dependent representation model, semantic priming corresponds to textbase repetition without word form repetition. For example it is conceivable that if a person speaks about *gardening*, the receiver will have an expectation about certain words (plant, tree, seed, fertilizing), and it might be conceived that this semantic expectation is the cause of the decrease in fixation time as the text progresses. However, we can discard this possibility as the sole factor governing our observations, because this would have predicted that the repetition of lemmas (as an extreme case of semantic similarity) would lead to the same result. Instead, we observed that a decrease in duration required repetition of the precise word form.

In our view, the most parsimonious explanation of our results is to consider the repetition effect, at least partially, as a consequence of increased predictability, as given by the context. Within this frame, predictability could be thought of as an experimental quantification of the strength of the situation model, as postulated in Raney's model (Raney, 2003). Repeated words are predictable within a thematic context, and predictability is expected to be lower when two instances occur in different thematic contexts. In this framework, the reduction in gaze durations on successive appearances is not only produced by the presence of previous instances, but also by the strength of the context. Undoubtedly, repetition enhances the effect of the context.

Our results suggest that repetition led to an increase in predictability. However, without a direct measure of predictability for longer texts, one can only speculate about the relation between this variable and repetition in long-text reading. These results on multiple word repetition can only be observed during long-text reading, and so it comes with a cost. One of the difficulties of conducting these studies is that the corpus is much larger and, hence, computing cloze predictability for each individual word would require great effort. Because predictability is a very relevant implicit factor in this study, one could conceive of different possibilities to overcome this limitation in future studies. Several alternatives have been proposed to estimate predictability computationally: transition probabilities (Frisson et al., 2005; Keller & Lapata, 2003; McDonald & Shillcock, 2003), latent semantic analysis (Landauer & Dumais, 1997; Ong & Kliegl, 2008; Pynte et al., 2008a), and surprisal (Boston et al., 2008, 2011; Demberg & Keller, 2008). As far as we know, when including these measures in LMMs along with cloze predictability and frequency, they accounted for a significant amount of unique variance in fixation durations but were related more closely to frequency than to cloze predictability (Ong & Kliegl, 2008). Indeed, none of these alternatives rendered cloze predictability irrelevant. In this respect, the repetition number may move us a step closer. The need for such a measure is evident from its use in computational models of control of eye movement in reading. This is one of the first studies to investigate the dependence of fixation durations on multiple covariates during reading in natural situations in which sentences were not isolated but were embedded in a natural text.

Acknowledgments

We thank Yamila Sevilla, Daniel M. Low, Carolina Gattei, and Gerardo Fernández for their thoughtful reading. We thank Thomas A. Gavin, Professor Emeritus, Cornell University, for help with editing the English in this article. This work is funded by the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), M.S. is sponsored by the James McDonnell Foundation 21st Century Science Initiative in Understanding Human Cognition Scholar Award.

Supplemental material

Supplemental material, data, and R scripts for statistical analyses and the generation of figures are available at http://reading.liaa.dc.uba.ar/.

References

Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology*, *17*, 364–390.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2013). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi:10.18637/jss.v067.i01

Bodner, G. E., & Masson, M. E. (2001). Prime validity affects masked repetition priming: Evidence for an episodic resource account of priming. *Journal of Memory and Language*, *45*, 616–647.

Boston, M. F., Hale, J., Kliegl, R., Patil, U., & Vasishth, S. (2008). Parsing costs as predictors of reading difficulty: An evaluation using the potsdam sentence corpus. *Journal of Eye Movement Research*, *2*, 1–12.

Boston, M. F., Hale, J. T., Vasishth, S., & Kliegl, R. (2011). Parallelism and syntactic processes in reading difficulty. *Language and Cognitive Processes*, *26*, 301–349.

Bowers, J. S. (2000). In defense of abstractionist theories of repetition priming and word identification. *Psychonomic Bulletin & Review*, *7*, 83–99.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.
Carr, T. H., Brown, J. S., & Charalambous, A. (1989). Repetition and reading: Perceptual encoding mechanisms are very abstract but not very interactive. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 763.
Chamberland, C., Saint-Aubin, J., & Legere, M. A. (2013). The impact of text repetition on content and function words during reading: further evidence from eye movements. *Canadian Journal of Experimental Psychology*, *67*, 94–99.
Clifton, C., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, and R. L. Hill (Eds.), *Eye movements: A window on mind and brain*, (pp. 341–371). New York, NY: Elsevier.
Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The eyelink toolbox: eye tracking with matlab and the psychophysics toolbox. *Behavior Research Methods, Instruments, & Computers*, *34*, 613–617.
Dambacher, M., Kliegl, R., Hofmann, M., & Jacobs, A. M. (2006). Frequency and predictability effects on event-related potentials during reading. *Brain Research*, *1084*, 89–103.
Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, *109*, 193–210.
Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, *20*, 641–655.
Fernández, G., Shalom, D. E., Kliegl, R., & Sigman, M. (2014). Eye movements during reading proverbs and regular sentences: the incoming word predictability effect. *Language, Cognition and Neuroscience*, *29*, 260–273.
Findlay, J. M., & Walker, R. (1999). A model of saccade generation based on parallel processing and competitive inhibition. *Behavioral and Brain Sciences*, *22*, 661–674.
Frisson, S., Rayner, K., & Pickering, M. J. (2005). Effects of contextual predictability and transitional probability on eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 862.
Heller, D., & Müller, H. (1983). On the relationship of saccade size and fixation duration in reading. In R. Groner, C. Menz, D. F. Fisher, & R. A. Monty (Eds.), *Eye movements and psychological functions: International views* (pp. 287–302). Hillsdale, NJ: Erlbaum.
Hyönä, J., & Niemi, P. (1990). Eye movements during repeated reading of a text. *Acta Psychol (Amst)*, *73*, 259–280.
Keller, F., & Lapata, M. (2003). Using the web to obtain frequencies for unseen bigrams. *Computational Linguistics*, *29*, 459–484.
Kinoshita, S. (2006). Additive and interactive effects of word frequency and masked repetition in the lexical decision task. *Psychon Bull Rev*, *13*, 668–673.
Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: the influence of past, present, and future words on fixation durations. *J Exp Psychol Gen*, *135*, 12–35.
Kuperman, V., Dambacher, M., Nuthmann, A., & Kliegl, R. (2010). The effect of word position on eye-movements in sentence and paragraph reading. *Q J Exp Psychol A*, *63*, 1838–1857.
Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*, 211.
Legge, G. E., Pelli, D. G., Rubin, G. S., & Schleske, M. M. (1985). Psychophysics of reading. I. Normal vision. *Vision Research*, *25*, 239–252.
Levy, B. (2001). Text processing: Memory representations mediate fluent reading. In M. Neveh-Benjamin, M. Moscovitch, and H. L. Roediger, III (Eds.), *Perspectives on human memory and cognitive aging: Essays in honour of Fergus Craik* (pp. 83–97). Philadelphia, PA: Psychology Press.
Levy, B. A., & Burns, K. I. (1990). Reprocessing text: Contributions from conceptually driven processes. *Canadian Journal of Psychology*, *44*, 465.
Logacev, P., & Vasishth, S. (2013). em2: A package for computing reading time measures for psycholinguistics. Retrieved from https://cran.r-project.org/src/contrib/Archive/em2/
Marslen-Wilson, W., Tyler, L. K., Waksler, R., & Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychological Review*, *101*, 3–33.
McDonald, S. A., & Shillcock, R. C. (2003). Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science*, *14*, 648–652.
Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, *106*, 226.
O'Brien, E. J., Raney, G. E., Albrecht, J. E., & Rayner, K. (1997). Processes involved in the resolution of explicit anaphors. *Discourse Processes*, *23*, 1–24.
Ong, J. K. Y., & Kliegl, R. (2008). Conditional co-occurrence probability acts like frequency in predicting fixation durations. *Journal of Eye Movement Research*, *2*, 3.
Padró, L., & Stanilovsky, E. (2012). Freeling 3.0: Towards wider multilinguality. In N. C. C. Chair, K. Choukri, T. Declerck, M. U. Dogan, B. Maegaard, J. Mariani, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the Language Resources and Evaluation Conference (LREC 2012). ELRA, European Language Resources Association (ELRA), Istanbul, Turkey*, pp. 2473–2479.
Pelli, D. G., Tillman, K. A., Freeman, J., Su, M., Berger, T. D., & Majaj, N. J. (2007). Crowding and eccentricity determine reading rate. *Journal of Vision*, *7*, 20.

Pollatsek, A., Rayner, K., & Balota, D. A. (1986). Inferences about eye movement control from the perceptual span in reading. *Perception & Psychophysics*, *40*, 123–130.
Pynte, J., New, B., & Kennedy, A. (2008a). A multiple regression analysis of syntactic and semantic influences in reading normal text. *Journal of Eye Movement Research*, *2*, 4.
Pynte, J., New, B., & Kennedy, A. (2008b). On-line contextual influences during reading normal text: a multiple-regression analysis. *Vision Res*, *48*, 2172–2183.
Raney, G., Therriault, D., & Minkoff, S. (2000). Repetition effects from paraphrased text: Evidence for an integrated representation model of text representation. *Discourse Processes*, *29*, 61–81.
Raney, G. E. (2003). A context-dependent representation model for explaining text repetition effects. *Psychon Bull Rev*, *10*, 15–28.
Raney, G. E., & Rayner, K. (1995). Word frequency effects and eye movements during two readings of a text. *Can J Exp Psychol*, *49*, 151–172.
Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372.
Rayner, K. (2009). Eye movements in reading: Models and data. *Journal of Eye Movement Research*, *2*, 1.
Rayner, K., Raney, G. E., & Pollatsek, A. (1995). Eye movements and discourse processing. In R. Lorch & E. O'Brien (Eds.), *Sources of coherence in reading* (pp. 9–36). Hillsdale, NJ: Lawrence Erlbaum Associates.
Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review*, *3*, 504–509.
Sakamoto, Y., Ishiguro, M., & Kitagawa, G. (1986). *Akaike information criterion statistics*. Dordrecht, The Netherlands: D. Reidel.
Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461–464.
Sebastián-Gallés, N., Martí, M. A., Cuetos, F., & Carreiras, M. (1998). LEXESP: Léxico informatizado del español. Ediciones de la Universidad de Barcelona, Barcelona.
Singer, M., & Halldorson, M. (1996). Constructing and validating motive bridging inferences. *Cognitive Psychology*, *30*, 1–38.
Sternberg, S. (1969). The discovery of processing stages: Extensions of donders' method. *Acta Psychologica*, *30*, 276–315.
Taylor, W. L. (1953). Cloze procedure: A new tool for measuring readability. *Journalism and Mass Communication Quarterly*, *30*(4), 415–433.
Tenpenny, P. L. (1995). Abstractionist versus episodic theories of repetition priming and word identification. *Psychonomic Bulletin & Review*, *2*, 339–363.
Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.
Van Petten, C., Kutas, M., Kluender, R., Mitchiner, M., & McIsaac, H. (1991). Fractionating the word repetition effect with event-related potentials. *Journal of Cognitive Neuroscience*, *3*, 131–150.
Vanyukov, P. M., Warren, T., Wheeler, M. E., & Reichle, E. D. (2012). The emergence of frequency effects in eye movements. *Cognition*, *123*, 185–189.
von der Malsburg, T., & Angele, B. (2016). False positives and other statistical errors in standard analyses of eye movements in reading. arXiv preprint. arXiv: 1504.06896.